

**What is claimed is:**

- 1           1.     A method of generating acoustic morphemes, comprising:  
2                   receiving training speech;  
3                   selecting candidate phone-phrases from the training speech;  
4                   selecting salient phone-phrases from the candidate phone-phrases  
5     based on salience measurements;  
6                   clustering the salient phone-phrases based on semantic and  
7     syntactic similarities into acoustic morphemes.
- 1           2.     The method of claim 1, further comprising:  
2                   storing the acoustic morphemes in a database.
- 1           3.     The method of claim 2, wherein the acoustic morpheme database  
2     is used by a speech recognition and understanding system.
- 1           4.     The method of claim 1, wherein the step of selecting candidate  
2     phone-phrases includes:  
3                   filtering the training speech;  
4                   selecting all observed phone sequences of a predetermined length;  
5     and  
6                   selecting as candidate phone-phrases the phone sequences that  
7     are of at least the predetermined length.
- 1           5.     The method of claim 1, wherein the training speech includes at  
2     least one of verbal and non-verbal speech.
- 1           6.     The method of claim 5, wherein the non-verbal speech includes the  
2     use of at least one of gestures, body movements, head movements, non-  
3     responses, text, keyboard entries, keypad entries, mouse clicks, DTMF codes,  
4     pointers, stylus, cable set-top box entries, graphical user interface entries and  
5     touchscreen entries.
- 1           7.     The method of claim 1, wherein the training speech includes  
2     multimodal forms.
- 1           8.     The method of claim 1, wherein the training speech is  
2     untranscribed.
- 1           9.     The method of claim 1, wherein the training speech is transcribed.

1           10.    The method of claim 1, wherein the salient phone-phrases are  
2 selected using a test for significance.

1           11.    The method of claim 1, wherein the salient phone-phrases are  
2 clustered into acoustic morphemes using a distortion measure between the  
3 salient phone-phrases.

1           12.    The method of claim 11, wherein the distortion measure is based  
2 on at least one of string distortion, semantic distortion and syntactic distortion.

1           13.    A method of generating morphemes, comprising:  
2                   receiving training speech;  
3                   selecting candidate sub-morphemes from the training speech;  
4                   selecting salient sub-morphemes from the candidate sub-  
5 morphemes based on salience measurements;  
6                   clustering the salient sub-morphemes based on semantic and  
7 syntactic similarities into morphemes.

1           14.    The method of claim 13, wherein the morphemes are at least one  
2 of acoustic and non-acoustic.

1           15.    The method of claim 13, further comprising:  
2                   storing the morphemes in a database.

1           16.    The method of claim 15, wherein the morpheme database is used  
2 by a speech recognition and understanding system.

1           17.    The method of claim 13, wherein the step of selecting candidate  
2 sub-morphemes includes:

3                   filtering the training speech;  
4                   selecting all observed sub-morpheme sequences of a  
5 predetermined length; and  
6                   selecting as candidate sub-morphemes the sub-morpheme  
7 sequences that are of at least the predetermined length.

1           18.    The method of claim 13, wherein the training speech includes at  
2 least one of verbal and non-verbal speech.

1           19.    The method of claim 18, wherein the non-verbal speech includes  
2 the use of at least one of gestures, body movements, head movements, non-

3 responses, text, keyboard entries, keypad entries, mouse clicks, DTMF codes,  
4 pointers, stylus, cable set-top box entries, graphical user interface entries and  
5 touchscreen entries.

1 20. The method of claim 13, wherein the training speech includes  
2 multimodal forms.

1 21. The method of claim 13, wherein the training speech is  
2 untranscribed.

1 22. The method of claim 13, wherein the training speech is transcribed.

1 23. The method of claim 13, wherein the salient sub-morphemes are  
2 selected using a test for significance.

1 24. The method of claim 13, wherein the salient sub-morphemes are  
2 clustered into morphemes using a distortion measure between the salient sub-  
3 morphemes.

1 25. The method of claim 24, wherein the distortion measure is based  
2 on at least one of string distortion, semantic distortion and syntactic distortion.